2011

Film Review Aggregators and Their Effect on Sustained Box Office Performance

Nicholas Krishnamurthy
Claremont McKenna College

CLAREMONT McKENNA COLLEGE

FILM REVIEW AGGREGATORS AND THEIR EFFECT ON SUSTAINED BOX OFFICE PERFORMANCE

SUBMITTED TO

PROFESSOR DARREN FILSON

AND

DEAN GREGORY HESS

BY

NICHOLAS KRISHNAMURTHY

FOR

SENIOR THESIS

FALL / 2011

November 28, 2011

Acknowledgements

I would like to thank my parents for their constant support of my academic and career endeavors, my brother for his advice throughout college, and my friends for always helping to keep things in perspective. I would also like to thank Professor Filson for his help and support during the development and execution of this thesis.

Abstract

This thesis will discuss the emerging influence of film review aggregators and their effect on the changing landscape for reviews in the film industry. Specifically, this study will look at the top 150 domestic grossing films of 2010 to empirically study the effects of two specific review aggregators. A time-delayed approach to regression analysis is used to measure the influencing effects of these aggregators in the long run. Subsequently, other factors crucial to predicting film success are also analyzed in the context of sustained earnings.

Table of Contents

1. Introduction

Movies are highly lucrative products whose success is nearly impossible to predict. Their role and affect on American culture is undeniable, but the makers of widely released films, the few large studios and the many smaller production companies, seem to have limited ability to predict the success of a given film.

Intuitively, this lack of predictability makes sense. A film is, in fact, a one-of-a-kind product. Each filmmaking company, therefore, never releases the same product twice, nor knows what to expect upon conceiving the idea for each product. This has not, however, stopped these companies from attempting to predict the success of their future film endeavors (Simonoff and Sparrow 2000).

Meanwhile, film also exists in culture as an art form, and, like any other art form, is constantly consumed, critiqued, and criticized (Ginsburgh and Weyers 1999). These two worlds, for the most part, exist independently of each other. That is, a studio would not care if a film is considered high art if it makes a considerable amount of money at the box office. On the other hand, an art critic cares very little about how much a movie makes, and often might prefer a movie more willing to challenge the typical formula for popular appeal and take risks in its filmmaking technique. This type of critic has a preference for what could be referred to as "high art." The two worlds collide, however, in the case where artistic merit influences consumer preference.

Holbrook (2005) discusses whether consumers' real preference actually mirrors that of these critics. Do consumer's take value out of viewing high art? Or do consumers take value out of enjoying what others *claim* as high art? Independent of the motivation, film companies have long sought an answer to this question – does quality, as defined by film critics, affect their overall revenue?

Critical reviews play an important role in the consumption of film because, similar to other products such as books, theater, and restaurants, the quality of the product is hard to judge prior to consumption (Boatwright, Basuroy, and Kamakura 2007). Reviews themselves can come in a number of different forms. Classically, a select few high-profile critics have had a significant monopoly on the marketplace for critical reviews for such products. Boatwright, Basuroy, and Kamakura (2007) note that Frank Rich and Clive Barnes, theater reviewers for the *New York Times* and *New York Post* respectively, had a much greater effect on the success of a theater show than did critics from other newspapers. The same effect can be seen to some extent in the film industry. The film industry is differentiated, however, as a more populist medium – seeing a movie is less costly, more convenient, and more easily accessible – and therefore, while film reviews do play an important role in the industry, they must be viewed in a specific light.

Consumer preference for film differs from that of other art forms due to consumers' lower standards and desire for escapist entertainment. Although tough to numerically judge, one might safely assume that a high proportion of consumers do not see a specific film based on its artistic appeal. That is, most consumers see a movie in hopes of being entertained, humored, scared, and to escape from or connect with other facets of daily life. Such a consumer must not, and would not neglect all reviews as analyses of artistic merit, but may reduce the level at which they judge a film as worthy of their time, money, and therefore viewership. For example, a possible viewer who has seen trailers, advertisements, and posters for an upcoming blockbuster epic based on a best-selling novel, may in fact choose to see the movie even if a specific critic gave the film 5 out of 10 stars, but may not choose to see the movie if it received 2 out of 10 stars from the same critic. It is safe to assume that if an expert critic were to rate a film 5 out of 10 stars, then the film, at least in the mind of that specific critic, is not worthy of artistic praise. The consumer, however, chooses to see the film for a variety of other reasons independent of artistic praise. In light of this analysis of consumer preference, the forms for consuming critical reviews have begun to change.

Rather than relying on one specific high-profile critic, consumers may prefer the option of viewing a collection of critical reviews on the movie. Even further, the consumers may be less interested in the reviews themselves than a numerical analysis of the collection of scores assigned to a film. One-hundred moderately positive reviews may have more influence on a specific consumer than one glowing review from a high-profile critic praising the film's artistic merit.

An aggregated approach to collecting critical reviews for film may not, and likely will not, predict the entire financial success of a film. A high budget movie will still likely open to strong numbers at the box office. As the run of a film continues, however, and word of mouth spreads, an aggregated approach to film reviews may have a significant influencing affect, both positive and negative, on consumers' desire to see a film. This study will look at multiple aggregated approaches to film criticism, as well as other factors previously studied by academics (Basuroy, Chatterjee, and Ravid 2003; De Vany and Walls 1999; De Vany and Walls 2002; King 2007; Ravid 1999; Simonoff and Sparrow 2000; Terry, Butler, and De'Armond 2005) and empirically analyze their effect, over time, on a film's financial success.

2. Literature Review

Much research has been done on the topic of predicting box office success. Litman (1983) explains how film production and marketing consists of three major factors that all play a crucial role in how a film is received by critics as well as its audience. The first of these factors is that of the "creative sphere." This consists of the story and script on which the eventual film will be based. In this context, films can garner success from genre, style, or known creative entities. That is, an epic action-adventure movie, a well-written independent drama, and a film based on a best-selling novel could all succeed while harnessing different creative ambitions. The second factor, the production budget, plays a large role in determining how well the filmmakers can execute the original creative vision. The production budget influences a film's ability to afford a star actor or actress, large explosions, lavish sets, and a number of other items that can influence a film's quality. Lastly, a film's rating, assigned by the Motion Picture Association of America (MPAA), plays an important role in influencing how a studio or production company will market and how the audience will receive the film. These ratings, G, PG, PG-13, R, or NC-17, can narrow or widen a film's reach, with negative and positive possibilities for both outcomes. A film's marketing budget often reflects a combination of these three factors. For example, star power, genre, and rating all have direct impact on how much the distributor will market a specific film (Prag and Casavant 1994).

The film review itself can play a number of important roles in connection with how audiences consume their entertainment. Film reviews can influence how an audience perceives the quality of a film, they can create a reputation for certain talent associated with a film, and they can serve as free advertising (Basuroy, Chatterjee, and Ravid 2003). The entertainment industry itself stands out as having an extremely high level of cultural impact, and therefore film reviews have a disproportionate affect on the typical American consumer (Eliashberg, Elberse, and Leenders 2006). Audience reactions to these reviews are best evaluated as an inverted-U shaped model. As audiences learn more about a film through reviews and advertising they become curious, and therefore more likely to go see the film. Once they reach a certain information peek, however, they begin to feel oversaturated with information, and become less intrigued with the film and therefore less likely to attend (Wyatt and Badger 1990).

Basuroy, Chatterjee, and Ravid (2003) outline three issues crucial to the understanding of film reviews and their effect on the performance of a film. First, contrary to intuition, a film review actually gains power over time as studios begin to leverage good reviews in order to overpower any negative reviews. A studio, for example, will begin to quote specific positive reviews in their advertising. Secondly, negative reviews have a larger effect on a film's performance than positive reviews. This is mostly a result of negativity bias, but in relation to the first issue, the negative reviews will loose power as time progresses. Thirdly, studios may use stars or high production budgets to counteract any possible negative reviews since viewers are more likely to disregard reviews in light of these "high profile" factors.

Reviewers themselves can behave in multiple fashions. Holbrook (1999) explains how reviewers can act as either expert critics or journalistic reviewers. The expert critics concern themselves with providing a subjective critique of the film's artistic merit. Such a review has little relation to what an audience may enjoy, but instead exists as a high-brow judgment on the film as an art form rather than a consumptive good. The journalistic reviewer attempts to recommend certain films in accordance to what they think the consumer will enjoy. Therefore, rather than making judgments on artistic merit, they are acting as advisors for the general populous in terms of recommending films. Many studies have been conducted analyzing the effect of general reviews on box office performance, and, on the whole, the journalistic reviewers tend to correlate more closely with box office performance than do the expert critics.

Multiple studies have been conducted analyzing the predicting and influencing powers critics hold over box office performance. The predicting power speculates that critics do not convince consumers to see the film, but rather predict how many viewers *will* see the film, and therefore predict how a film will perform at the box office. A conflicting approach, that of the influencing power of critics, theorizes that critics have the power to actually affect the box office performance of a film. In terms of measurement, an influencing critic should have significant effect on performance immediately upon release of the review (likely tied with release of the film), whereas a predicting critic should correlate more closely with the late performance of a film, since the review itself has little effect on performance following release. Alternatively, a reviewer could exhibit qualities of both an influencer and predictor, and correlate closely with the entire release of the film (Basuroy, Chatterjee, and Ravid 2003). Eliashberg and Shugan's (1997) empirical study tests the potential influencer and predictor hypotheses, and concludes that, because critic reviews are highly correlated with late box office performance but have a low correlation with performance within the first four weeks of release, their data supports the predictor rather than the influencer hypothesis.

There has also been a considerate amount of research focusing on the philosophical and psychological components to the reviews and consumption of certain art forms (of which film can be considered one). Ginsburgh and Weyers (1999) discuss the evolution of this thought since Plato, and land on only three consensus viewpoints. First, that expert critics should be responsible for the assessment of artistic merit rather then the public. Second, that any review of artistic merit is of a subjective nature, and that it is impossible to objectively critique art. Lastly, artistic merit and popular appeal are inevitably blended upon release of an artistic work, and only time can separate "good art" from fashion.

The resulting question from Ginsburgh and Weyer's (1999) explanation is why do consumer preferences tend to differ from the experts' analysis of artistic merit. Holbrook's (2005) study offers two possible explanations for why consumers tend to behave differently. His first hypothesis, titled "The Dignity-of-the-Common-Person Hypothesis," suggests that consumers' evaluation of artistic merit *does* fall in line with that of expert critics. It suggests, however, that consumers also act upon preferences independent of artistic merit. An example is consumer preference towards escapist films which sometimes demonstrate little artistic ability but allow the consumer to submit their mind entirely to the entertainment. The second hypothesis, titled "The What-a Wonderful-World Hypothesis," suggests that consumer preferences exist entirely independent of expert judgments of good and bad art. Consumers may in fact have entirely different opinions on what is "good" than do the expert critics. Although the Holbrook (2005) study had its own self-noted limitations, there did appear significant support for the Dignity-of-the-Common-People hypothesis.

Reviews also differ in terms of the scale at which they are presented to a consumer. A review could come in the form of a celebrity critic, such as Robert Ebert, known for his reviews in the form of newspaper editorials, television programs, and most recently, blogs (Ebert 2011). Reviews could be collected from top sources around the United States and compiled into a table in the Sunday newspaper. A source, often online, could also collect many reviews from multiple sources and provide summary statistics assessing the general consensus – such is the case with Rotten Tomatoes (www.rottentomatoes.com) and Metacritic (www.metacritic.com). Boatwright, Basuroy, and Kamakura's (2007) study analyzes reviews in terms of both the individual and the aggregate approach. They note, first of all, that the intuitive power of the aggregate outweighs that of the individual in most cases. That is, if you disregard the level of expertise of the specific critics, a consensus opinion will always prove more powerful. Extrapolating further, depending on specific consumer preference, there arises a point for each consumer at which the aggregate opinion, independent on the level of expertise, may outweigh a few specific critics of high expertise. The authors argue further that, in an industry with such a high cultural significance, the expert reviewers maintain a level of fame that allows for heightened influence over the consumer. There maintains a balance, therefore, between the individual and the aggregate in terms of which has a stronger effect on consumers and therefore on box office performance.

The effects of reviews tend to change and adapt over the course of a film's run at the box office. De Vany and Walls (1996) analyze the information flow of reviews over time using Bose-Einstein dynamics. Summarized, they analyze the snowball effect in response to both positive and negative reviews. That is, consumers and reviewers are likely to spread the word to their friends and peers about both positive and negative reactions to certain films. As the word spreads, that opinion begins to dominate the marketplace. Some studios will actually adapt their marketing strategy to reflect the information flow. If they notice that a film is faring particularly well in certain areas and poorly in others, they may focus the remainder of their marketing on the more receptive areas (Simonoff and Sparrow 2000). These theories of information flow for film reviews also relate back to the discussion of aggregate review statistics – the aggregate statistics may be more likely to affect word of mouth appraisal than the opinion of one critic, whose particular viewpoint may be more likely to appeal to the artistic preferences of a single consumer.

Significant discussion has been focused on the topic of uncertain attributes in the film industry. Since each film is, in itself, unique and uncertain, the process of developing a film often undertakes some level of signaling. Those involved with the film will take actions that have a secondary impact of informing consumers of the film's quality. Prag and Casavant (1994) note that the production budget itself is a form of signaling. A high production budget may increase the quality of the film by allowing for sets, costumes, effects, and more, but it also signals to the public that the filmmakers believe in the film's quality and expect it to perform well and recoup the production expenses. Ravid's (1999) study of star power and its affect on box office gross also recognizes the power of signaling, focusing on the decision making power of studio executives. Ravid (1999) argues that the public, aware of the risk each studio executive takes upon green-lighting a film and attaching stars and directors, has less skepticism for high-profile and high-risk films. If the executive is willing to risk their job, the film must be somewhat decent.

Ravid's (1999) study, referenced previously, performs a close analysis of the role of stars in relation to film performance at the box office. The study looks at 175 films release between 1991 and 1993, and regresses on a number of factors relating to genre, production budget, and rating. To measure star power Ravid (1999) uses a complex system looking at both Academy Award (Oscar) wins and nominations. Ravid (1999) concludes that star power does have some affect in driving film revenue. The study, however, notes the "rent-capture hypothesis," which claims that a star actor, actress, or director may capture any additional revenue garnered from their contribution to the film. Since most talent contracts create for some incentive system based on overall film performance, the stars will continue to rake in profits as the film continues to perform well at the box office.

Many empirical analyses of star-power, being innately hard to measure due to the secrecy of talent contracts and the constantly fluctuating demand and consumer preference for specific talent, often rely on Academy Awards wins and nominations to compile numeric data. Ravid's (1999) study, which took such an approach, concludes that award wins and nominations do have some impact on performance.

De Vany and Walls (2002) look at risk in relation to a film's MPAA rating. They note that R-rated movies are innately more risky because their subject matter is often geared to an older and narrower audience. A studio can increase its potential upside by making less R-rated movies, and focusing more on movies geared towards kids. Many movies that are G or PG-rated, however, are animated. Many studios have already taken action and formed animation units, hired filmmakers and animators keen to the new technology associated with high-budget animated products, and begun releasing multiple animated films each year. Statistically, however, the data is unclear in whether it supports this risk analysis. Although R-rated movies do have a lower median revenue rate, that could actually reflect lower risk, since they often cost less to produce and are not expected to earn as high of revenue as other films. Simonoff and Sparrow (2000) also note that an animated film takes considerably longer to produce, and requires a larger budget, often due to the unique set of personnel required for such complex productions.

De Vany and Walls' (1999) study focuses on the distribution of film revenue in relation to a film's inherent risk. The distribution of films is heavily skewed, with many performing poorly or moderately, and only a few garnering large profits. Those that do succeed, however, make enough profit to keep the industry highly lucrative. De Vany and Walls (1999) note, however, that due to the distribution there exists no clear average for film revenue, nor a clear standard deviation. While this does not and has not stopped researchers from attempting to predict factors leading to box office success, it does pose an issue to any such model.

Terry, Butler, and De'Armond (2005) also look at a combination of possibilities in their regression analysis of factors effecting box office performance, focusing on a sample of all movies from 2001 through 2003 that either opened in 25 theaters in the United States or eventually grew to 100 theaters at some point in its run, leading to a sample of 505 films. The study looks at the simpler measure of awards per film, concluding that each award nomination is worth more than 6 million dollars. The increased earnings arise out of a signaling effect, with each nomination reflecting general industry acclaim based on expert judgments. The study also looks at sequels and genre as factors that could have an affect on revenue. Sequels are concluded to have a positive affect, while certain genres show mixed results – action movies correlate positively with revenue, while children's movies, surprisingly, correlate negatively. Finally, the study looks at Rotten Tomatoes scores in attempt to gauge critical appeal. They conclude that, while looking only at overall revenue rather than revenue as it changes over time, high critical review scores do positively correlate with revenue, noting a few important exceptions.

King (2007) uses Metacritic as the basis for his analysis of critical reviews and their effect on box office earnings. Metacritic aggregates review scores from a number of sources, but, unlike Rotten Tomatoes, assigns each review a score for the film out of 10, and averages them all to give a Metacritic score more reflective of total quality, rather than Rotten Tomatoes' "good enough" score. King's (2007) analysis looks at films in 2003, and concludes that, while looking at total box office revenue, high critic scores have a weak but positive affect on film revenue, but that it is often mediated by limitations associated with wide and limited releases. King (2007) notes that for wide releases, critic scores make less difference, but that for films with limited releases but high critic scores, the praise is still unlikely to propel the film to a wide release.

3. Methods

Previous studies have also used data from film aggregators in their approach to analyzing how multiple variables effect the box office performance of a film, but these studies were conducted towards the beginning of the twenty-first century (King 2007; Terry, Butler, and De'Armond 2005). In order to more closely analyze the effects of film aggregators in the industry's current state, this study will look at films released in 2010. There are a number of online sources that collect the top films from each year and rank them by their box office performance. Some difficulty occurs when choosing how many movies to include in a study like this. Many movies are released each year – some reach millions of viewers and others likely do not reach one-hundred.

Holbrook (2005) uses the website www.worldwideboxoffice.com (accessed September 19, 2011-November 28, 2011) to collect all films released in the year 2000. Doing the same for the year 2010 results in the collection of 428 films. The highest grossing of these is *Toy Story 3*, which grossed $1,063,171,911 worldwide. The lowest grossing is *Detention*, which failed to break $500 worldwide. Obviously such a range of financial success would be inappropriate for a study analyzing how critical reviews affect box office performance. Even further, films such as *Detention*, likely released to only one theater for a small audience, would have a low probability of being reviewed by any critic. It is therefore necessary to make some revenue cutoff at which any film below will be unhelpful to this study. Further, many websites, such as www.worldwideboxoffice.com, look closely at the total revenue of a film, both domestic and international. The online film aggregators used in this study, however, maintain most of their prominence in the United States, and are unlikely to have a significant influencing effect on international revenue. It is therefore more useful to study domestic revenue instead of worldwide revenue. This should not be mistaken, however, with international films, on whose domestic performance the film aggregators may have an influencing effect. We therefore reject methods used previously by academics that look at "all" films released in a given year, or that look at worldwide revenue.

Instead, the website Box Office Mojo (www.boxofficemojo.com, accessed September 19, 2011-November 28, 2011) was used to compile a list of the top 150 grossing films of 2010, ranked by their domestic revenue. Of these, the highest grossing was still *Toy Story 3*, with domestic revenue of $415,004,880, and the lowest grossing was *Mao's Last Dancer*, with domestic revenue of $4,817,770. All the information for domestic, international, and weekly revenue was available from Box Office Mojo, as well as the studio responsible for the film, the open and close dates, and the production budget.

One of the websites this study will analyze, Metacritic (www.metacritic.com, accessed September 19, 2011-November 28, 2011), aggregates the reviews from approximately 40 national sources. The score assigned from each source, independent of its rating system, is then converted to a 100 point scale. For example 1 ½ thumbs up out of 2 would equal 75 out of 100 on the Metacritic scale. Metacritic also contains some subjective bias, weighting what it deems thorough or professional reviews ahead of brief, less articulate reviews. Then, based on the film's overall Metacritic score, that is, the average of all the individual scores, the film is termed to have "universal acclaim" (81-100), "generally favorable reviews" (60-80), "mixed or average reviews" (40-59), "generally unfavorable reviews" (20-39), or "overwhelming dislike" (0-19). The overall score is presented clearly at the top of the webpage for each movie, and Metacritic associates a color, red, yellow, or green, for bad, moderate, or good respectively, with the film for visual stimulation. Metacritic also provides every review counted, with their score and a brief synopsis of the review, if the consumers wish to look further into the film. Finally, Metacritic allows consumers who have seen the movie to report their score, and subsequently compiles a "user score." While this may be an influencing factor on some consumers, it does not represent critical reviews or their influence on consumer's willingness to see a film. It also represents a biased selection of consumers, since only those consumers with strong preferences for or against the film are likely to submit their scores. For these reasons user score will be ignored for this study.

The second source used for in this study, Rotten Tomatoes (www.rottentomatoes.com, accessed September 19, 2011-November 28, 2011), uses similar methods to aggregate reviews, but provides a more holistic approach to presenting its results. Rather than collecting reviews from a limited number of respected sources, Rotten Tomatoes widens its reach and can often include reviews from over 200 sources. It then subcategorizes its most respected critics as "top critics." Rotten Tomatoes, like Metacritic, collects and presents the scaled ratings from each review, but that is not the main score presented. Instead, Rotten Tomatoes presents the percentage of reviews that were positive. An aggregate percentage of over 60 is deemed "fresh" and awarded the logo of a fresh tomato. An aggregate percentage of less than 60 is deemed "rotten" and awarded the logo of a squished green tomato. A film that has a percentage over 75 *and* over 40 counted reviews is deemed "certified fresh" and awarded the logo of the Rotten Tomatoes seal. The site also provides each review for consumers who wish to look further, and briefly summarizes the general consensus on the film if enough information has been collected. Finally, like Metacritic, Rotten Tomatoes allows users to rate the film, but this will be mostly ignored for the purposes of this paper.

In order to both further explain the factors relating to a film's success, as well as to more closely understand the relation of film aggregators on financial performance, multiple other variable were included. These were compiled from a number of sources, and are as follows:

1. The *Production budget* of the film. This plays an important role in a regression analysis of film revenue, as a high production budget could outweigh the abundance of negative reviews for a film. Also, it would make intuitive sense for the production budget to correlate with the opening gross of a film, since a high production budget may signal higher quality or at least the expectation of consumer appeal (Prag and Casavant 1994). Production budgets, however, are often confidential and not released to the public. Nevertheless, multiple sources give reasonable estimates for most movies. Box Office Mojo lists production budgets for most movies. For movies without their production budget listed on Box Office Mojo, other reputable sources, such as International Movie Database (www.imdb.com, accessed September 19, 2011-November 28, 2011) were used.
2. *Number of total screens* on which a film was shown. In many cases a film will be released initially at fewer theaters to generate press, and will widen its release over time. Figures accounting for screens upon a film's initial release will often under-represent the width of a film's ultimate release.
3. Whether a film is a *sequel*. This binary variable denotes sequels and prequels to previous films. The variable does not, however, count remakes or spin-offs as sequels. For example, neither the 2010 remake, *The Karate Kid*, with Jackie Chan, nor *Get Him to the Greek*, the spin-off film from 2008's *Forgetting Sarah Marshal*, were counted as sequels.

4. *Genre indicators* denote the genre of a specific film. Genre is broken down into family, drama, and action/comedy. The genre of a film was determined from the International Movie Database. For films that covered multiple genres, the first listed genre was used. For example, the comedy/drama *Due Date* is listed as a comedy.
5. The film *rating*, as chosen by the Motion Picture Association of America. The rating for a film is either G (general audiences), PG (parental guidance suggested), PG-13 (parents strongly cautioned), or R (restricted).

Most previous studies on the subject of predicting financial success of films have used total revenue, in some form, as their dependent variable (Basuroy, Chatterjee, and Ravid 2003; De Vany and Walls 1999; De Vany and Walls 2002; King 2007; Ravid 1999; Simonoff and Sparrow 2000; Terry, Butler, and De'Armond 2005). Theoretically, however, review aggregators should have a delayed influencing effect on the success of a film. Many movies open to a strong box office, independent of their quality. For example, the 2010 film *The Last Airbender* opened domestically to over $40 million, but received aggregated scores of 6 and 20 from Rotten Tomatoes and Metacritic respectively. This idea of a delayed influencer effect contrasts Eliashberg and Shugan (1997), who claim that any influencer effect should be seen immediately upon release of the film. Of more interest to this specific study, however, is whether the review aggregators have an effect on a film's continued performance, or performance in the stage in a film's run where one may speculate that quality, rather than spectacle or other factors, has a large effect on consumer preference.

In order to analyze the delayed effect of film review aggregators on financial performance, the domestic revenue for the first 2 weeks of release is used as an independent variable. The dependent variable, therefore, is the domestic revenue post-week-3. In this case we assume that the aggregated review scores do not have a significant effect on the first two weeks of revenue, and that consumers who wish to see a movie of poor quality will do so because of previous expectations, knowledge, or excitement about the movie, and are therefore likely to see it within the first two weeks. Instead, we will look at the influencing effect of review aggregators on the continued performance of a film, during a period when consumers are likely more driven by the quality of a film.

4. Analysis

In our analysis of film review aggregators and their effect on sustained box office performance we will individually analyze the effect of both Metacritic and Rotten Tomatoes scores. Subsequently, other factors will be added to the regression analysis to determine whether they too have an effect on revenue. Since Metacritic and Rotten Tomatoes scores both explain the same outcome, we cannot analyze both simultaneously, and thus must use two separate models when analyzing the data.

As seen from Table 1, The Rotten Tomatoes scores have a mean of 51, but have a rather high standard deviation of close to 27. This means that there are a considerable amount of films that score either extremely low or extremely high on the Rotten Tomatoes scale.

Metacritic scores appear more condensed. Their mean of 53 reflects similar conditions as the Rotten Tomatoes scores – that most films are of average quality. The Metacritic standard deviation of just above 17, however, is considerably lower than for Rotten Tomatoes. This reflects the nature of the two ratings. Rotten Tomatoes reflects a binary response of good or bad

Table 1. Summary statistics for Rotten Tomatoes and Metacritic scores

	Mean	Std. Dev.	Min	Max	Median
Rotten Tomatoes	51.22	26.94	4	99	51
Metacritic	53.03	17.37	18	95	53

Figure 1. Distribution of Rotten Tomatoes and Metacritic ratings

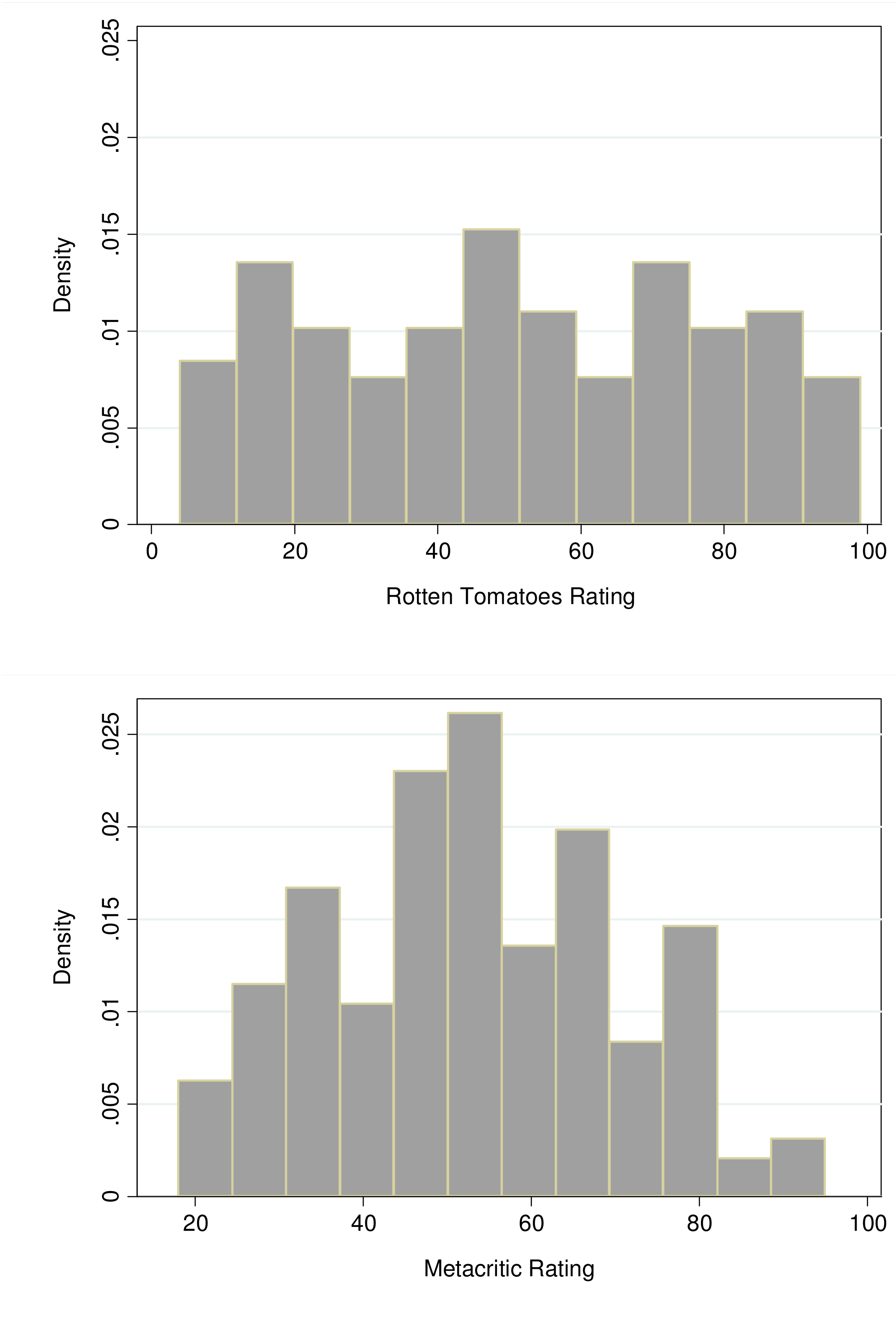
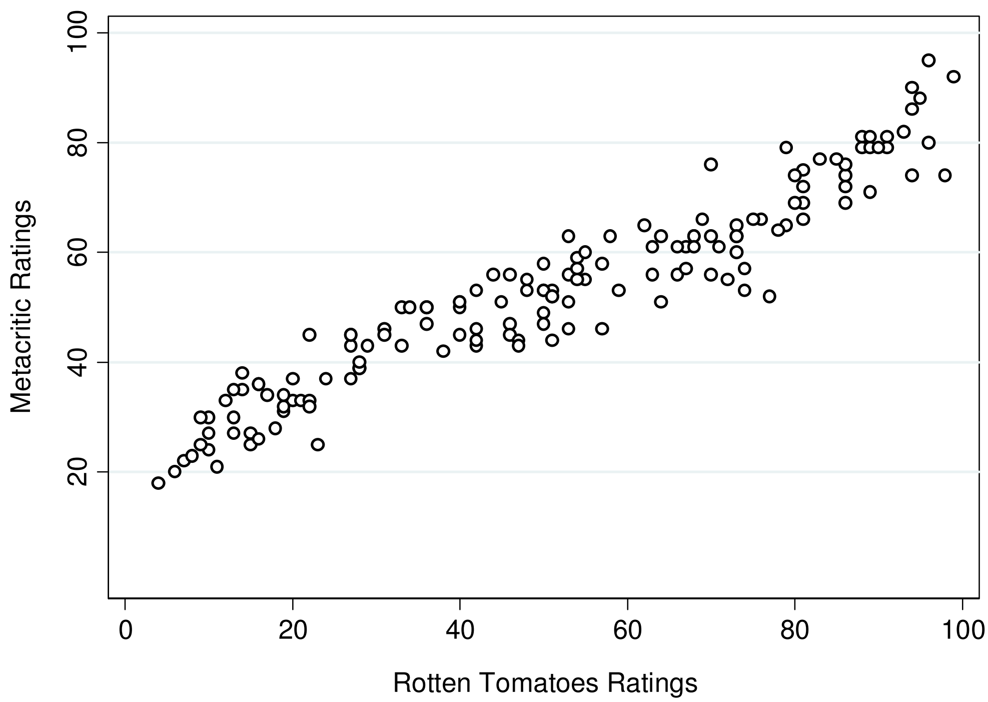


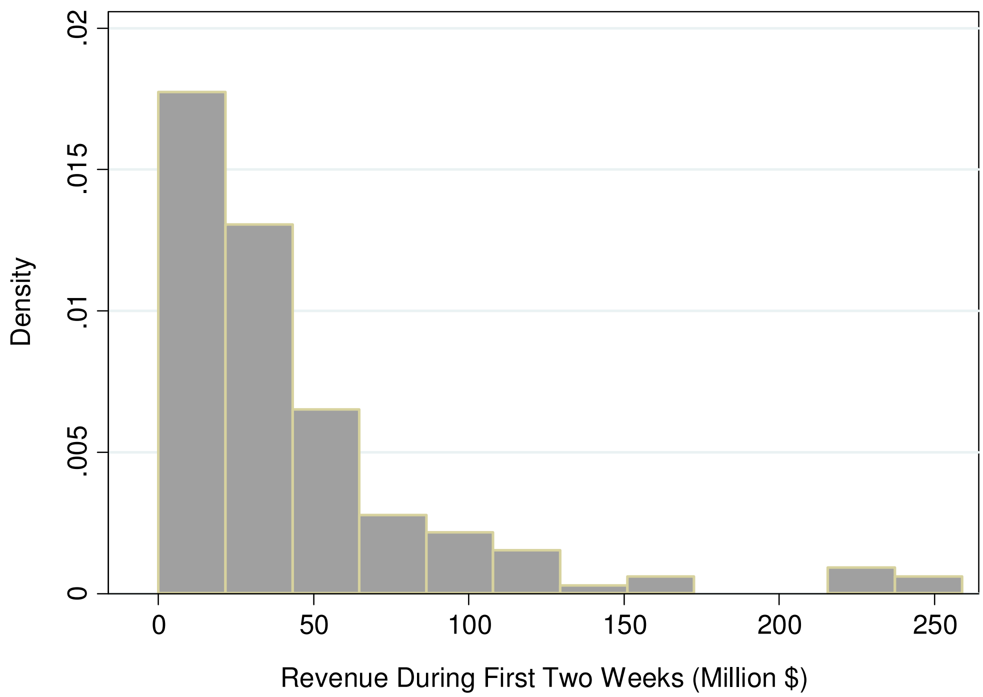
Figure 2. Scatter plot of Metacritic vs. Rotten Tomatoes Ratings



while Metacritic reflects each reviewers score on a gradient scale. For example we can look at the 2010 film *The Bounty Hunter* which received a Rotten Tomatoes score of 7 but a Metacritic score of 22. This means, in theory, that 93 percent of critics thought the film was bad, but that the average grade for those critics was somewhere near 20/100. Thus one can see how Metacritic scores, by their nature, will have less variance than those of Rotten Tomatoes.

A comparison of the Rotten Tomatoes and Metacritic scores for each movie proves useful in understanding how the two variables relate to each other. Figure 2 plots both the Metacritic ratings and the Rotten Tomatoes ratings for each of the 150 films. Even a brief glance at the data clearly shows a high correlation. Also apparent, as discussed previously, is that the Metacritic ratings do not drop much below 20, while multiple films, such as *Vampires Suck*, *The Last*

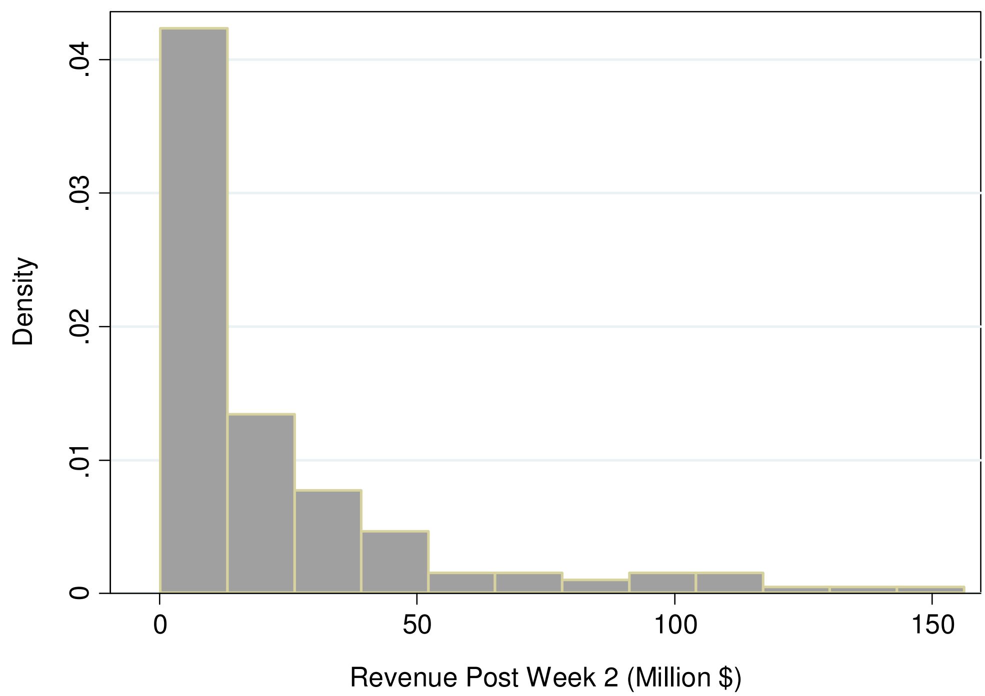
Figure 3. Density chart for revenue during the first two weeks of release



Airbender, *The Bounty Hunter*, and *Furry Vengeance*, have Rotten Tomatoes ratings below 10. The graph, along with the difference in variances of the two ratings, shows that the vertical intercept of any Metacritic vs. Rotten Tomatoes model will be greater than 0.

The distribution of revenue, unlike that of critic scores, does not follow a centralized normal distribution. Figure 3 shows that the distribution of revenues in the first two weeks has a long right tail implying a positive skew. The density graph shows that only a select few films grossed over 100 million dollars in their first two weeks, but some films had two-week revenues of over 200 million. Revenues gained after the first two weeks of release, as displayed in Figure 4, show the same skew to an either further extent. No films were able to garner domestic revenue of more than 160 million after the first two weeks of release, and the majority of films

Figure 4. Density chart for revenue during weeks 3 and on

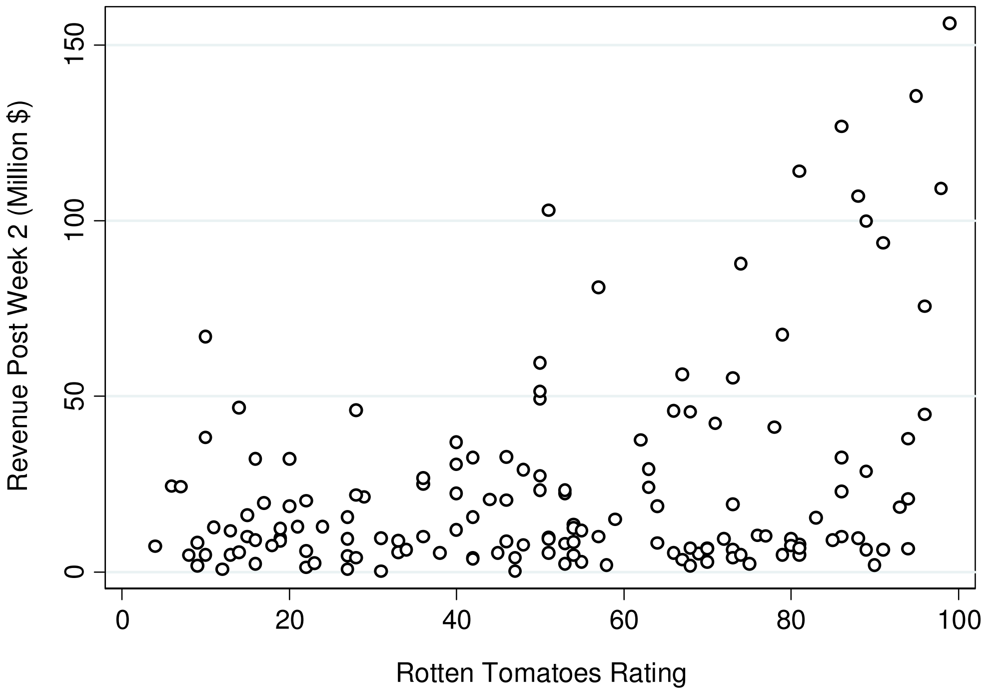


earned less than 50 million in their extended run.

Due to the positive skew of both revenue during the first two weeks and revenue after the first two weeks, we can use a logarithmic transformation to more accurately analyze the data in the regression. The result is effectively to normalize the data, but it will have an impact in the final analysis, causing the regression results to be reflected as percentages rather than as increments.

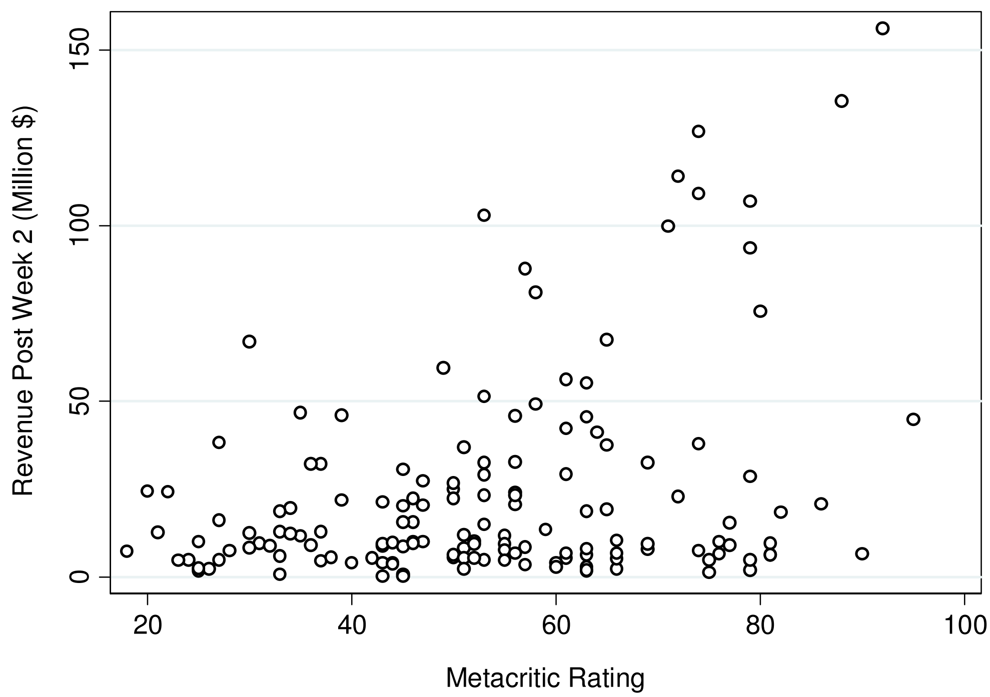
The direct relation between both the review aggregators and revenue after the second week of release can be seen in Figure 5 and Figure 6 for Rotten Tomatoes and Metacritic respectfully. The graphs both show an upward curve, implying that, at least on the surface, a higher Rotten Tomatoes or Metacritic score results in increased revenue post-week-2.

Figure 5. Scatter plot of sustained revenue vs. Rotten Tomatoes rating



The main regression uses the log of revenue in the first two weeks and either Rotten Tomatoes or Metacritic scores as regressors, and uses revenue post week 2 of release as the dependent variable. Due to problems associated with limited releases, some films from the top 150 films of 2010 were unable to be included in the regression. Many films begin their theatrical run by only showing on a select few screens, often at premiere cinemas in either Los Angeles or New York. After a few weeks at such theaters, they then expand slowly to a wider national audience. The trouble occurs when discussing how to calculate the first two weeks of revenue. Using the actual revenue from the first two weeks would understate the initial appeal, but beginning the measurement upon wide release would ignore multiple weeks of strong performance, even if on a limited basis. With no suitable way to handle the delayed effect for

Figure 6. Scatter plot of sustained revenue vs. Metacritic rating



these such films simultaneously with that of normal, wide, releases, the select films were removed for regressions involving time-sensitive revenue data. The resulting 127 films were used for the regression analysis.

The results for the base Rotten Tomatoes regression are displayed in regression 1 of Table 2. Not surprisingly, the revenue in the first two weeks correlates highly with the revenue post-week-2 – a good open indicated higher revenue in a film's extended run. More precisely, the data indicates that, holding everything else constant, a 1 percent increase in revenue during the first two weeks results in a 1.2 percent increase in revenue in subsequent weeks. The coefficient for Rotten Tomatoes scores is also both positive and statistically significant. The coefficient implies that, holding everything else constant, a one unit increase in the Rotten

Table 2. Base and expanded regressions using Rotten Tomatoes scores

Dependent Variable: Revenue Post Week 2 of Release

Regressor	**(1)**	**(2)**	**(3)**
Rotten Tomatoes Rating (X_1)	.0064*** (0.0023)	.0070*** (0.0023)	.0057*** (.0021)
Log of Revenue in First 2 Weeks of Release (X_2)	1.157*** (0.107)	1.090*** (0.130)	1.279*** (.172)
Log of the Production Budget (X_3)		0.103 (0.070)	.130* (.077)
Binary Variable for an MPAA Rating of G or PG (X_4)		.425*** (0.107)	.438*** (.161)
Number of Theaters (X_5)			.000 (.000)
Binary Variable for a Sequel (X_6)			-.081 (.180)
Binary Variable for the Family Genre (X_7)			.190 (.177)
Binary Variable for the Drama Genre (X_8)			.197 (.159)
Intercept	-4.063** (1.848)	-4.844** (2.039)	-7.686*** (2.781)
Summar Statistics			
R^2	0.7352	0.7609	.7722

Heteroskedasticity robust standard errors are given in parenthesis under coefficients. The individuals coefficient is statistically significant at the * 10%, ** 5%, or *** 1% significance level using a two sided test.

Tomatoes score results in a .64 percent increase in revenue for weeks 3 and on. For example, using the mean post-week-2-revenue, a 1 unit increase in Rotten Tomatoes score would result in an increase in revenue of close to $190,000.

The same regression, with Metacritic scores substituted for Rotten Tomatoes scores, displayed in regression 1 of Table 3, shows similar results. The coefficient for the first two weeks of revenue is statistically significant and a near 1-to-1 relationship with the dependent variable. The Metacritic coefficient, like the Rotten Tomatoes coefficient, is positive and statistically significant. The Metacritic coefficient, however, is higher, and implies that, holding everything else constant, a one unit increase in Metacritic score results in a 1.24 percent increase in revenue. Once again using the mean post-week-2-revenue as an example, a one unit increase in Metacritic score would result in a $360,000 increase in revenue. Also, since the standard deviation of Metacritic is lower than that of Rotten Tomatoes, a one unit increase has a greater influence on revenue-post-week-2 than it would for Rotten Tomatoes. Both these regressions, although not accounting for other possible factors, show that Rotten Tomatoes and Metacritic scores both have a statistically significant positive influence on sustained box office revenue.

In order to analyze the effect of other possible factors on sustained revenue, variables for production budget, theaters, sequels, genre, and MPAA rating were all added to the regression. Production budget was transformed into a logarithmic variable using the same approach described earlier for both revenue variables. Production budget, similar to film earnings, has a positive skew. This implies that, although a select few films have a very high production budget, most films have smaller budgets. Thus, using a logged production budget variable allows us to treat the variable as normally distributed, and analyze the result using percentages.

Table 3. Base and expanded regressions using Metacritic scores

Dependent Variable: Revenue Post Week 2 of Release

Regressor	**(1)**	**(2)**	**(3)**
Rotten Tomatoes Rating (X_1)	.0124*** (0.0037)	.0132*** (0.0038)	.0106*** (.0035)
Log of Revenue in First 2 Weeks of Release (X_2)	1.151*** (0.065)	1.091*** (0.130)	1.254*** (.174)
Log of the Production Budget (X_3)		0.092 (0.071)	.116 (.077)
Binary Variable for an MPAA Rating of G or PG (X_4)		.433*** (0.106)	.438*** (.157)
Number of Theaters (X_5)			.000 (.000)
Binary Variable for a Sequel (X_6)			-.065 (.181)
Binary Variable for the Family Genre (X_7)			.170 (.172)
Binary Variable for the Drama Genre (X_8)			.153 (.160)
Intercept	-4.281** (1.777)	-5.000** (1.954)	-7.377*** (2.781)
Summar Statistics			
R^2	0.7419	0.7675	.7750

Heteroskedasticity robust standard errors are given in parenthesis under coefficients. The individuals coefficient is statistically significant at the * 10%, ** 5%, or *** 1% significance level using a two sided test.

Multiple regressions were conducted using a variety of combinations of additional variables, a film aggregator, and revenue for the first two weeks as regressors. Some of the additional variables, independent of the regression, proved statistically insignificant to predicting sustained revenue. The binary variable for a sequel was negetively correlated but had a high enough variance that it proved insignificant. The theaters variable, oddly, had a coefficient of 0. This makes little intuitive sense, since the more theaters in which a film was shown, the more money you would expect a film to earn.

Genre was broken into two binary variables, drama and family. Family films are unique due to their wide audience appeal. Meanwhile, dramas appeal more to adults and, one could imagine, benefit from longer runs as patience and busy lives influence movie-going habits. Not included in these two variables, and therefore explained if both binary variables are left at 0, are action movies and comedies. Due to the similar appeal of both genres, and the frequent occurrence in the data of "action-comedies," the two genres were left as a single group.

The drama variable proved statistically insignificant throughout every regression. Contrastingly, the family indicator proved positive and significant for many regressions – specifically any regression for which MPAA rating indicators were not used. MPAA ratings were described using a G/PG indicator for films appealing to all audiences, and leaving PG-13 and R films (there were no NC-17 films in the data set) explained by a G/PG value of 0. The coefficient for G/PG was both positively correlated and statistically significant in every regression. Further, however, for every regression that included both the G/PG indicator and the family indicator, the family indicator proved to be statistically insignificant. The apparent positive correlation between family movies and sustained performance, therefore, is better explained by a correlation between low MPAA ratings and sustained performance. Holding

MPAA ratings constant, the genre of a movie has little to no effect on its sustained box office performance.

The production budget had varied levels of statistical significance depending on the specific regression and the other variables included. It was, however, low enough that its inclusion in any expansion from the base regression seems necessary. The second, expanded regressions, therefore, as seen in regression 2 of Table 2 and Table 3 for Rotten Tomatoes and Metacritic respectfully, include the aggregator score, the log of revenue in the first two weeks, the production budget, and the G/PG indicator. The coefficients to the aggregators stay rather consistent, increasing by a little less than a tenth of a percent for both Rotten Tomatoes and Metacritic. The coefficients for the revenue through week 2 also stayed consistent, decreasing by a little less than a tenth of a percent for each regression. For both regressions the logged production budget had a p-value of approximately 20 percent, too high to prove significance, but not high enough to rule out completely. The coefficient itself claims that, holding the other variables constant, a 1 percent increase in production budget would result in approximately a 10 percent increase in revenue. Most shocking, however, is the statistically significant coefficient to the G/PG indicator. The coefficient explains that, holding the other variables constant, a G or PG film will make over 40 percent more at the box office after week 2 than a film rated PG-13 or R. The regressions using Rotten Tomatoes and Metacritic also have high R-squared values indicating that, respectfully, 76.1 and 76.8 percent of the variation in revenue post week 2 of release can be explained by the aggregator rating, revenue during the first two weeks, production budget, and MPAA rating.

The mixed results for the production budget, theaters, sequels, genre, and MPAA rating that were discussed earlier all relate to their effect on sustained revenue. Other studies have shown that these same variables, however, have significance in the short run (De Vany and Walls 2002; King 2007; Prag and Casavant 1994; Ravid 1999; Simonoff and Sparrow 2000; Terry, Butler, and De'Armond 2005). In order to confirm that the year 2010 is not an outlier, and that the insignificance of the factors listed above is only in relation to sustained revenue, similar regressions were run with revenue in the first two weeks as the dependent variable. These regressions show that production budget, theaters, and sequels all have a positive effect in the short run. The result for theaters makes sense, as most of the capacity concerns for theaters are in the first few weeks. Thus, after week 2, the number of theaters in which a film is shown makes less of a difference. The significance of sequels also makes sense, as consumers are often more invested in sequels, and therefore more likely to see them in the first two weeks. Genre and MPAA rating, however, which previously indicated the strength of family oriented low-rating films, become insignificant in the new regression. Hence, family films are unlikely to draw most of their audience in the first two weeks and benefit from an extended run. Ultimately, when regressed on revenue during the first two weeks of release, the factors which showed little effect in our first regressions do prove significant.

Table 4. All variables regressed on revenue during the first two weeks of release

Dependent Variable: Revenue During the First Two Weeks

Regressor	**(1)**	**(2)**
Rotten Tomatoes Rating (X_1)	.0044*** (.0015)	
Metacritic Rating (X_2)		.0085*** (.0024)
Log of the Production Budget (X_3)	.130** (.050)	.116** (.049)
Binary Variable for an MPAA Rating of G or PG (X_4)	-.167 (.111)	-.162 .109
Number of Theaters (X_5)	.001*** (.000)	.001*** (.000)
Binary Variable for a Sequel (X_6)	.263** (.102)	.270*** (.100)
Binary Variable for the Family Genre (X_7)	-.027 (.158)	-.045 (.155)
Binary Variable for the Drama Genre (X_8)	.019 (.764)	-.018 (.102)
Intercept	12.114*** (.764)	12.105*** (.737)
Summar Statistics		
R^2	.7711	.7765

Heteroskedasticity robust standard errors are given in parenthesis under coefficients. The individuals coefficient is statistically significant at the * 10%, ** 5%, or *** 1% significance level using a two sided test.

5. Discussion

The results of the regression analysis performed in this study indicate that review aggregators do have a significant effect on the performance of a film during its extended run at the domestic box office. Of the variables studied, the variables for Rotten Tomatoes and Metacritic ratings proved to be among the most influential and statistically significant.

The logarithmic form allows us to more easily handle the highly skewed data for domestic revenue, both during and after the second week of release. The result from the regressions may seem small, since, as we stated earlier, a one unit increase in score for Rotten Tomatoes and Metacritic would lead to, using the average domestic gross post-week-2, a $190,000 and $360,000 increase, respectfully, in sustained revenue. The implications, however, are that if a studio or production company were to invest further in quality, potentially raising their Rotten Tomatoes or Metacritic score by a significant amount, they could potentially make millions in extra revenue. Having said that, a film often takes its shape early and it may be difficult to, on a limited budget, greatly increase the quality of the film. In that case, the study suggests that an investment in quality, from the beginning and during the production process, could greatly benefit all parties involved.

Of surprise, and arguably the most influential of all the variables studied, was the binary indicator of whether a movie has a rating of either G or PG from the MPAA. While the enormity of its effect does come as a surprise, the effect itself makes intuitive sense. Movies geared towards younger audiences have a wide appeal – often kids, parents, grandparents, and all ages in between are motivated to attend a certain movie, for example, *Toy Story 3* – and could maintain higher levels of enthusiasm, especially amongst kids, over longer periods of time, compared to other blockbuster films. Observational analysis shows that the main studios have already begun a shift towards producing more high profile family films, often of the animated variety, in hopes of capturing this effect (De Vany and Walls 2002; Simonoff and Sparrow 2000).

Contrary to studies performed previously (King 2007; Ravid 1999), the regressions performed in this study did not find statistical significance for the production budget having any effect on sustained domestic revenue. Of note, however, is that previous studies have found that production budget has an effect on *total* revenue, rather than sustained revenue. The effects of production budget found in this study, however, while statistically insignificant, do shed light onto a possible positive correlation between budget and sustained revenue.

Further variables, such as theaters, sequels, and genre all provided little insight into what factors effect sustained domestic revenue over a films extended run. The empirical analysis concludes, therefore, that review aggregators have a small yet important role in influencing a movie's performance after its second week of release domestically.

6. Conclusion

It is tough to precisely discuss the role of critics and critical reviews on a film's performance at the box office. Eliashberg and Shugan (1997) claim that, in order for a critical review to influence a film's performance, the effect must be seen immediately upon release of the film. While not refuting that point, this study attempts to claim that review aggregators have in fact changed the way people access, consume, and analyze reviews, and that they have, as a result, changed the way consumers decide whether or not to see a movie.

Review aggregators, specifically Rotten Tomatoes and Metacritic, do not attempt to provide only the best and most scholarly, professional, or artistic reviews. The overall effect is, instead, to compile the shared opinions of all those who job or hobby has led them to write a review. Some may be bad, and some may be good, but the result is a closer representation of traditional public appeal – some mix of artistic critique with an analysis of entertainment value.

The review is not released overnight or in the morning newspaper, and therefore the effect is not necessarily seen upon release of the film. Instead, as more reviews are added to the website, and word of mouth spreads, a film can see a long-run effect due, in part, to a positive score on Rotten Tomatoes or Metacritic.

The fact remains that success of a film is nearly impossible to predict. Higher budgets and an industry trend towards high-profile family friendly films are, however, examples of studios and production companies attempting to beat, or at least better the odds of producing a financially successful film. This study suggests that, while other factors cannot be ignored, the changing landscape of reviews, their increasingly accessible nature, and convergence towards general public opinion, all make review aggregators, and therefore film quality, an important factor in predicting a film's financial success, and particularly the success of a film during its extended run at the domestic box office.

7. Appendix

Table 5. List of 2010 films in alphabetical order

127 Hours
A Nightmare on Elm Street (2010)
Alice in Wonderland (2010)
Alpha and Omega
Babies
Black Swan
Blue Valentine
Brooklyn's Finest
Burlesque
Case 39
Cats & Dogs: The Revenge of Kitty Galore
Charlie St. Cloud
City Island
Clash of the Titans (2010)
Conviction
Cop Out
Country Strong
Cyrus
Date Night
Daybreakers
Dear John
Death at a Funeral (2010)
Despicable Me
Devil
Diary of a Wimpy Kid
Dinner for Schmucks
Due Date
Easy A
Eat Pray Love
Edge of Darkness
Extraordinary Measures
Fair Game (2010)
Faster
For Colored Girls
From Paris with Love
Furry Vengeance
Get Him to the Greek
Get Low
Going the Distance
Green Zone
Grown Ups
Gulliver's Travels
Harry Potter and the Deathly Hallows Part 1
Hereafter
Hot Tub Time Machine
How Do You Know
How to Train Your Dragon
Hubble 3D
I Am Love
Inception
Iron Man 2
It's Kind of a Funny Story
Jackass 3-D
Jonah Hex
Just Wright
Kick-Ass
Killers
Knight & Day
Leap Year
Legend of the Guardians: The Owls of Ga'Hoole
Legion (2010)
Let Me In
Letters to Juliet
Life as We Know It
Little Fockers
Lottery Ticket
Love and Other Drugs
MacGruber
Machete
Mao's Last Dancer
Marmaduke
Megamind
Morning Glory
My Soul to Take
Nanny McPhee Returns
Oceans
Our Family Wedding
Paranormal Activity 2

Percy Jackson & The Olympians: The Lightning Thief
Piranha 3D
Predators
Prince of Persia: The Sands of Time
Ramona and Beezus
Red
Remember Me
Repo Men
Resident Evil: Afterlife
Robin Hood
Salt
Saw 3D
Scott Pilgrim vs. the World
Secretariat
Sex and the City 2
She's Out of My League
Shrek Forever After
Shutter Island
Skyline
Splice
Step Up 3-D
Takers
Tangled
The American
The A-Team
The Back-Up Plan
The Book of Eli
The Bounty Hunter
The Chronicles of Narnia: The Voyage of the Dawn Treader
The Crazies
The Expendables
The Fighter
The Ghost Writer
The Girl Who Kicked the Hornet's Nest
The Girl Who Played with Fire
The Girl with the Dragon Tattoo
The Karate Kid
The Kids Are All Right
The King's Speech
The Last Airbender
The Last Exorcism
The Last Song
The Last Station
The Losers
The Next Three Days
The Other Guys
The Secret in Their Eyes
The Social Network
The Sorcerer's Apprentice
The Spy Next Door
The Switch
The Tourist
The Town
The Twilight Saga: Eclipse
The Warrior's Way
The Wolfman
Tooth Fairy
Toy Story 3
Tron Legacy
True Grit
Tyler Perry's Why Did I Get Married Too?
Unstoppable
Valentine's Day
Vampires Suck
Waiting for "Superman"
Wall Street: Money Never Sleeps
When in Rome
Winter's Bone
Yogi Bear
You Again
Youth in Revolt

References

Basuroy, Suman, Subimal Chatterjee, and Abraham Ravid. 2003. How critical are critical reviews? The box office effects of film critics, star power, and budgets. *Journal of Marketing* 67: 103-17.

Boatwright, Peter, Suman Basuroy, and Wagner Kamakura. 2007. Reviewing the reviewers: The impact of individual film critics on box office performance. *Quantitative Marketing and Economics* 5: 401-25.

De Vany, Arthur, and W. David Walls. 1999. Uncertainty in the movie industry: does star power reduce the terror of the box office? *Journal of Cultural Economics* 23: 285-318.

De Vany, Arthur, and W. David Walls. 1996. Bose-Einstein dynamics and adaptive contracting in the motion picture industry. *The Economic Journal* 106, no. 439: 1493-514.

De Vany, Arthur, and W. David Walls. 2002. Does Hollywood make too many R-rated movies? Risk, stochastic dominance, and the illusion of expectation. *The Journal of Business* 75, no. 3: 425-51.

Ebert, Roger. 2011. REVIEW --- Film Criticism Is Dying? Not Online --- Thanks to the Internet, there is more and better writing about movies than ever before, says Roger Ebert. *Wall Street Journal*, January 22.

Eliashberg, Jehoshua, Anita Elberse, and Mark A.A.M. Leenders. 2006. The motion picture industry: critical issues in practice, current research, and new research directions. *Marketing Science* 25, no. 6: 638-61.

Eliashberg, Jehoshua, and Steven M. Shugan. 1997. Film critics: influencers or predictors? *Jouurnal of Marketing* 61: 68-78.

Ginsburgh, Victor, and Sheila Weyers. 1999. On the perceived quality of movies. *Journal of Cultural Economics* 23: 269-83.

Holbrook, Morris B. 1999. Popular appeal versus expert judgments of motion pictures. *Journal of Consumer Research* 26: 144-55.

Holbrook, Morris B. 2005. The role of ordinary evaluations in the market for popular culture: do consumers have ‘good taste’? *Marketing Letters* 16, no. 2: 75-86.

King, Timothy. 2007. Does film criticism affect box office earnings? Evidence from movies released in the U.S. in 2003. *Journal of Cultural Economics* 31: 171-86.

Litman, Barry R. 1983. Predicting success of theatrical movies: an empirical study. *Journal of Popular Culture* 16, no. 4: 159-75.

Prag, Jay and James Casavant. 1994. An empirical study of the determinants of revenues and marketing expenditures in the motion picture industry. *Journal of Cultural Economics* 18: 217-35.

Ravid, Abraham S. 1999. Information, blockbusters, and stars: a study of the film industry. *Journal of Business* 72, no. 4: 463-92.

Simonoff, Jeffrey S., and Ilana R. Sparrow. 2000. Predicting movie grosses: winners and losers, blockbusters and sleepers. *Chance* 13, no. 3: 15-24.

Terry Neil, Michael Butler, and De'Arno De'Armond. 2005. The determinants of domestic box office performance in the motion picture industry. *Southwestern Economic Review* 32: 137-48.

Wyatt, Robert O., and David P. Badger. 1990. Effects of information and evaluation in film criticism. *Journalism and Mass Communication Quarterly* 67, no. 2: 359-68.